 INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA SÃO PAULO Campus Bragança Paulista	Nome:		Prontuário:	
	Curso: Análise e Desenvolvimento de Sistemas		Data Entrega: 31/05/2025	Período: Noturno
	Disciplina: Fundamentos da Ciência de Dados		Módulo: 5	Valor: 10,0
	Prof. César A. S. Lima, Dr.			Nota:
	Assinatura:		Turma:	

## Análise Exploratória com Statsmodel

Este estudo deverá ser realizado com o dataset público iris. Desenvolva seu estudo no jupyter notebook.

### Instrução:

O arquivo deverá ser salvo com o seguinte nome:

**Trabalho\_T2\_Nome\_Sobrenome.zip**

**Atenção:** Seu trabalho não será validado se o nome do arquivo não estiver identificado conforme instrução.

Faça a postagem do seu estudo, impreterivelmente, no **sábado dia 31/05/2025** no Moodle.

As bibliotecas a serem utilizadas no estudo serão as seguintes:

`import pandas as pd`

`import seaborn as sns`

`import matplotlib.pyplot as plt`

`import statsmodels.api as sm`

`from sklearn.datasets import load_iris`

O dataset iris será carregado segundo a linha de código a seguir:

`iris = load_iris()`


### ### 🎯 Problema de Negócio

Como as características morfológicas das flores de iris influenciam o comprimento da sépala 'sepal\_length' em função das variáveis independentes 'sepal\_width', 'petal\_length', 'petal\_width'.

### ### 📌 Contexto

O dataset Iris contém medidas de flores de três espécies diferentes de iris: setosa, versicolor e virginica. As variáveis incluem:

- Comprimento e largura da sépala
- Comprimento e largura da pétala
- Espécie da flor

 <p>INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA SÃO PAULO Campus Bragança Paulista</p>	Nome:		Prontuário:	
	Curso: Análise e Desenvolvimento de Sistemas		Data Entrega: 31/05/2025	Período: Noturno
	Disciplina: Fundamentos da Ciência de Dados		Módulo: 5	Valor: 10,0
	Prof. César A. S. Lima, Dr.			Nota:
	Assinatura:		Turma:	

### ### 🎯 Objetivo da Análise

Construir um modelo de regressão linear que permita:

- Entender quais variáveis têm maior impacto sobre o comprimento da sépala.
- Quantificar essa influência com base nos coeficientes do modelo.
- Avaliar a qualidade do modelo por meio de métricas estatísticas e testes de diagnóstico.

### ### 🔍 Metodologia

#### 1. \*\*Análise Exploratória\*\*:

- Estatísticas descritivas das variáveis.
- Matriz de correlação para identificar relações entre as variáveis.
- Gráficos de dispersão para visualizar as relações entre as variáveis.

#### 2. \*\*Regressão Linear\*\*:

- Ajuste de um modelo de regressão linear com a variável alvo sendo o comprimento da sépala.
- Avaliação dos coeficientes do modelo para entender a influência de cada variável.
- Testes de diagnóstico para verificar a normalidade dos resíduos.

### ### 📊 Resultados Esperados

- Identificação das variáveis que mais influenciam o comprimento da sépala.
- Modelo de regressão linear com boa capacidade preditiva ( $R^2$  elevado).
- Resíduos do modelo seguindo uma distribuição normal, sem evidência de heterocedasticidade ou autocorrelação.

### ### 📌 Conclusão

A análise exploratória e a regressão linear permitirão entender melhor as relações entre as características morfológicas das flores de íris e o comprimento da sépala.